Hearing Research 383 (2019) 107807

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Auditory streaming and bistability paradigm extended to a dynamic environment

Áine Byrne [a], John Rinzel [a, b], James Rankin [c, *]

[a] Center for Neural Science, New York University, 4 Washington Place, 10003, New York, NY, USA
[b] Courant Institute of Mathematical Sciences, New York University, 251 Mercer St, 10012, New York, NY, USA
[c] Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Rd, Exeter, EX4 4QF, UK

ARTICLE INFO

Article history:
Received 25 July 2019
Received in revised form
19 September 2019
Accepted 1 October 2019
Available online 5 October 2019

Keywords:
Auditory streaming
Bistability
Entrainment

ABSTRACT

We explore stream segregation with temporally modulated acoustic features using behavioral experiments and modelling. The auditory streaming paradigm in which alternating high- A and low-frequency tones B appear in a repeating ABA-pattern, has been shown to be perceptually bistable for extended presentations (order of minutes). For a fixed, repeating stimulus, perception spontaneously changes (switches) at random times, every 2–15 s, between an integrated interpretation with a galloping rhythm and segregated streams. Streaming in a natural auditory environment requires segregation of auditory objects with features that evolve over time. With the relatively idealized ABA-triplet paradigm, we explore perceptual switching in a non-static environment by considering slowly and periodically varying stimulus features. Our previously published model captures the dynamics of auditory bistability and predicts here how perceptual switches are entrained, tightly locked to the rising and falling phase of modulation. In psychoacoustic experiments we find that entrainment depends on both the period of modulation and the intrinsic switch characteristics of individual listeners. The extended auditory streaming paradigm with slowly modulated stimulus features presented here will be of significant interest for future imaging and neurophysiology experiments by reducing the need for subjective perceptual reports of ongoing perception.

1. Introduction

In a dynamic auditory world the brain must resolve ambiguity between sounds and isolate objects of interest. The segregation of distinct sound sources is a fundamental function of the auditory system. A valued paradigm for investigating such processes is auditory streaming, in which alternating high- A and low-frequency tones B appear in a repeating ABA-pattern (van Noorden, 1975). Initially heard as one integrated stream (Int), the probability of hearing two segregated streams (Seg) gradually builds up over several to tens of seconds (Anstis and Saida, 1985). Build-up occurs more rapidly with a large difference in tone frequency (DF) between A and B and at faster presentation rates. The first perceptual switch, typically from integrated to segregated, is followed by persistent alternations every 2–15 s between the two interpretations (auditory bistability) (Pressnitzer and Hupé, 2006; Kondo and Kashino, 2009; Winkler et al., 2012; Rankin et al., 2015; Denham et al., 2018). A striking aspect of auditory bistability (and perceptual bistability more generally (Rodríguez-Martínez and Castillo-Parra, 2018)) is that a fixed repeating stimulus gives rise to two or more perceptual interpretations. However, streaming in a natural auditory environment requires segregation of auditory objects with features that evolve over time.

Auditory bistability with the streaming paradigm provides a valued framework to explore the neural representations associated with different perceptual interpretations (Gutschalk et al., 2005) and the network of brain areas involved in perceptual switches (Schadwinkel and Gutschalk, 2011; Kondo and Kashino, 2009; Kashino and Kondo, 2012). Recent imaging studies have explored attentional effects with auditory bistability (Billig et al., 2018; Kondo et al., 2018). Neural responses to the streaming paradigm have been studied in primary auditory cortex (A1) of awake monkeys (Fishman et al., 2001, 2004; Micheyl et al., 2005; Knyazeva et al., 2018), in the forebrain of awake (Bee and Klump, 2004; Bee

* Corresponding author.
E-mail address: james.rankin@gmail.com (J. Rankin).

https://doi.org/10.1016/j.heares.2019.107807
0378-5955/

et al., 2010) and behaving (Itatani and Klump, 2014) songbirds, and in the auditory periphery (Pressnitzer et al., 2008) and A1 (Farley and Noreña, 2015) of anesthetized guinea pigs. The tonotopic organization of A1 and increased forward masking at higher presentation rates (Micheyl et al., 2005; Fishman et al., 2001, 2004) can explain the feature dependence of these responses. No study has claimed that the neural substrate for the perceptual state or switches in perceptual states lies in or before A1. Indeed, the only animal study with neural data recorded from behaving animals (Itatani and Klump, 2014) concluded that only stimulus features and not perceptual choice is encoded in songbird forebrain (analogous to A1).

Our recent study introduced the first neuromechanistic competition model of auditory bistability (Rankin et al., 2015), capturing the dynamics of alternations in a system of stochastic differential equations. It is assumed that competition downstream of A1 resolves ambiguous perception. The model uses dynamic inputs that directly link to sensory features as represented by the neuronal responses of pre-competition stages: inputs based on electrophysiologically-recorded primary auditory cortex (A1) responses to interleaved A and B tones (Micheyl et al., 2005). Units at the model's competition stage pool inputs from tonotopically-organized A1 and feature mechanisms commonly found in cortex: mutual inhibition, slow adaptation, noise and self-excitation on an NMDA-like timescale. This combination of mechanisms produces bistable alternations (intrinsic oscillations), capturing the dynamics of switches between integrated and segregated percepts (Pressnitzer and Hupé, 2006). Alongside new experiments it was shown that perception is bistable over a wide range of DF - values with long integrated (segregated) durations at low (high) DF and so-called *equidominance* between integrated and segregated at around DF = 5 st (similar mean percept durations for each percept) (Rankin et al., 2015). The work was recently extended to account for build-up (early bias towards integration) and the effects of stimulus interruptions and perturbations (Rankin et al., 2017).

In the present study, we explore auditory bistability in a non-static environment, by considering slowly and periodically varying stimulus features. Predictions generated from our computational model serve as a jumping off point for psychoascoustic experiments exploring the entrainment of perceptual alternations when a stimulus feature is periodically modulated with period T_{mod} (Fig. 1A). Here DF is sinusoidally modulated about a value of 5 semitones (st), which is close to equidominance with modulation turned off. With fixed DF inputs the model produces alternations between integration and segregation with mean duration T_{eq}. These intrinsic oscillations can interact with a periodically modulated stimulus to produce entrainment. With DF above 5 st, segregation is more likely and with DF below 5 st integration is more likely. For modulation with a period of $2T_{eq}$ we expected segregation for a duration T_{eq} in the half cycle with DF above 5 st and integration for a duration T_{eq} in the half cycle with DF below 5 st. For modulation periods near $2T_{eq}$ this pattern of two switches per modulation period can persist. Our modelling confirmed such expectations and led to the following modelling predictions (generated prior to our experiments): 1) alternations entrain to the stimulus modulation with two switches per modulation cycle, 2) switches into segregated (integrated) occur on the DF upswing (downswing), with most switches occurring before the maximal (minimum) DF values and 3) entrainment is stronger when the DF modulation period is near double the mean percept duration at equidominance ($T_{mod} = 2T_{eq}$). More generally this study address the following questions: Can perception entrain to a slowly varying stimulus in auditory streaming? What determines the strength of entrainment? Do the switching characteristics of individual listeners affect entrainment?

2. Materials and methods

2.1. Participants

Eighteen participants (10 female) with a mean age 23 years took part in the experiment and were reimbursed for their time at a $10 hourly rate. Procedures were in compliance with guidelines for research with human participants and approved by the University Committee on Activities Involving Human Participants at New York University (study IRB-FY2016-310). All participants provided written informed consent. Each participant completed three blocks, two 12 trial blocks and one 9 trial block, giving a total of 33 trials per participant. In the two 12 block trials, the participants completed 3 trials with no modulation before 9 trials where the modulation period was varied. The 9 trial block consisted of modulated trials only. In order to balance the number of switches across modulation periods, there were six $T_{mod} = 5$ s trials, nine $T_{mod} = 10$ s and twelve $T_{mod} = 20$ s. A 27×27 latin square design was used to determine the order of conditions for each participant. Trials were 3 min long, a choice which fit well with the block design, chosen modulation periods and constraints on session length. The blocks were run across two 90 min sessions on different days, one 12 trial block in the first session, the remaining two blocks in the second session. One participant who confused the response keys was excluded from the analysis.

2.2. Stimuli

The stimuli consist of repeating 125 ms pure tone ABA_ triplets where '_' indicates a silence also lasting 125 ms; each ABA_ triplet is 0.5 s in duration (similarly, 125 ms tones were used in Micheyl et al. (2005) and 120 ms tones were used in Pressnitzer and Hupé (2006)). The higher frequency B tones are a variable DF semitones (st) above the lower frequency A tones. Cosine squared ramps are used with 5 ms rise and fall times. During 3 min trials the tone sequence is played binaurally through etymotic headphones at 65 dB SPL. The DF was modulated around 5 st at a depth of 1.5 st (max DF 6.5 st, min DF 3.5 st). Previous studies (Rankin et al., 2015) showed equidominance for DF = 5 st, with mean dwell times of roughly 5 s, giving a mean full cycle (with one switch Int to Seg and one switch Seg to Int) of 10 s. As such, three different modulation periods were used, $T_{mod} = 5$, 10, 20 s, corresponding to half the natural perceptual cycle, the natural perceptual cycle and twice this value. Three different A tone base frequencies were used; 392 Hz, 494 Hz and 659 Hz, giving mean B tone frequencies (at DF 5 st) of: 523 Hz, 659 Hz and 880 Hz. Each tone frequency pair occurred an equal number of times for each modulation condition. A minimum 20 s interval between trials (and 40 s after the 4th and 8th trial for 12 block trials and after the 5th trial in the 9 trial block) was used after which participants could the next trial when ready. Roving of base frequencies and breaks between trials reduced the possibility for latent adaptation carrying over from one trial to the next.

2.3. Experimental procedures

Participants sat in an acoustically shielded chamber and indicated their perceptual responses with button presses on a keyboard. In a two alternative forced choice (2AFC) task participants were instructed to report the integrated percept when they heard the A and the B tones together in an alternating or galloping rhythm and the segregated percept when they heard two separate streams, one with only A tones and one with only B tones. The percepts were explained to the participants with auditory and visual illustrations to ensure that the participants understood the two interpretations and could clearly distinguish between them.

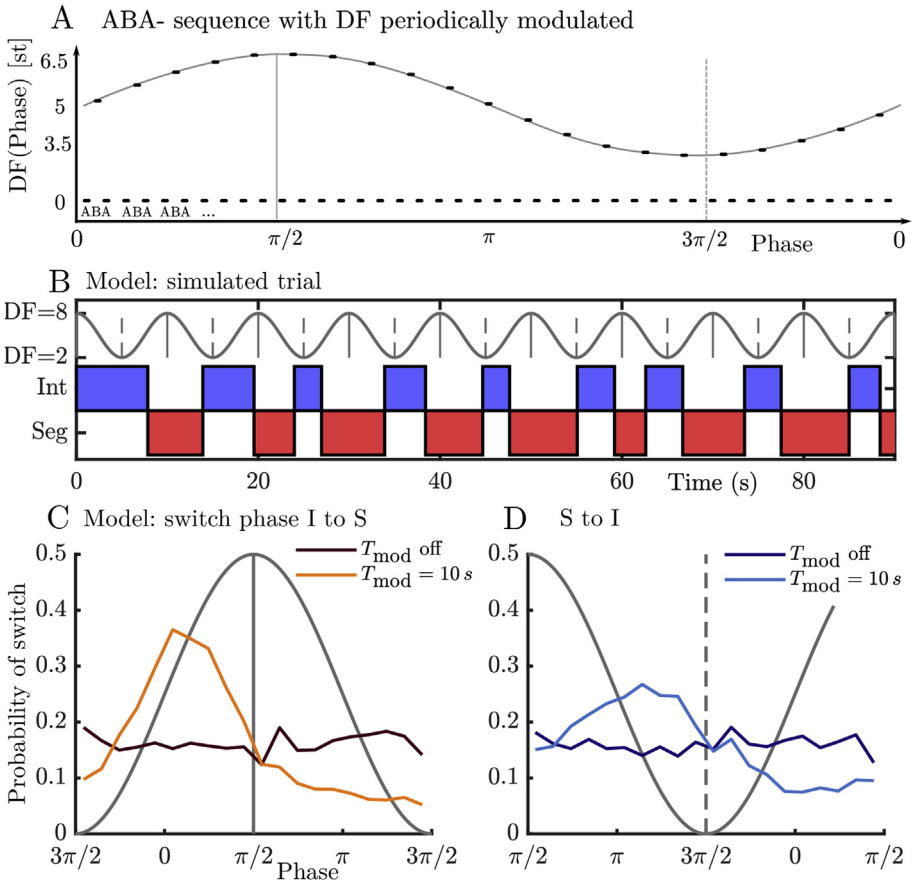

Fig. 1. Stimulus paradigm and model predictions. **A:** Auditory streaming paradigm with A and B tones separated by a time-dependent difference in tone frequency (DF). Tones are arranged in a repeating ABA-triplet pattern where B is a variable DF semitones (st) higher than that fixed A tones. One period of modulation shown for the case T_{mod} = 10 s where each triplet is 0.5 s in length. Stimulus is perceived as either one integrated stream (Int, more likely when DF is small) or two segregated streans (Seg, more likely when DF is large). **B:** Model simulation where DF is sinusoidally modulated ±3 st either side of the equidominance condition DF = 5 st. Perceptual responses entrain to the stimulus modulation with one switch from Int (blue) to Seg (red) and one switch from Seg to Int per period of modulation (grey). **C:** (**D**) Phase histograms were computed for switches in 320 simulated 4-min trials, a choice made before the conception of the block design for experiments, which used 3-min trials. Switches from Int to Seg (Seg to Int) are shown relative to the modulation of DF centered on the peak in modulation at Phase = $\pi/2$ (centered on the trough in modulation at Phase = $3\pi/2$). Switch timing is random with no modulation and when T_{mod} = 10 s locked to the rising phase of modulation for switches Int to Seg (**C**) and the falling phase of modulation for switches Seg to Int (**D**). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Participants were instructed to passively report their percepts without attempting to hear one perceptual organization over another. Participants reported their percepts by holding specific keys associated with each percept. The state of the two response buttons was recorded with a sampling rate of 100 Hz.

In this paper we considered bistability between integrated and segregated percepts for ABA-triplets, using a two-alternative forced choice (2AFC) task in our experiments. In studies where response keys are provided for integrated and segregated and participants are instructed to press neither key when their responses are "indeterminate" such responses are recorded for a very small fraction of presentation time (Pressnitzer and Hupé, 2006; Mill et al., 2013; Rankin et al., 2017). Here the task was 2AFC and no instruction for "indeterminate" responses was provided. Durations shorter than 0.5 s (one triplet) were excluded from the analysis. Given the 2AFC task, each percept duration was computed from the button press onset associated with one percept type up to the button press onset of the opposite percept type. The final (incomplete) duration was discarded for each trial.

2.4. Statistical analyses

All statistical analyses were carried out in the statistical package R. In the text throughout the manuscript, the Greenhouse-Geisser (G-G) corrected *p*-values are reported if a Mauchly sphericity (MS) test reached significance. In ANOVA tables the GG-corrected *p*-values are highlighted in bold if the MS test reached significance. In post-hoc analyses quoted *p*-values are Bonferroni corrected to account for multiple comparisons. Standard measures of effect size (generalized eta-squared η_G^2 and Cohen's *d*) are quoted for statistically significant results. ANOVA tables and posthoc analyses are reported in full in supplemental material Tables S1–S3.

In order to test the strength of phase entrainment, data were fit to von-Mises distributions varying parameters for the phase peak location (μ) and height (κ). The von-Mises distribution is a circular function defined with shape parameter $\kappa \geq 0$ centered at μ,

$$f(x;\mu,\kappa) = \exp(\kappa \cos(x-\mu)) \,/\, (2\pi I_0(\kappa)),$$

where $I_0(\kappa)$ is 0-order modified Bessel function. The function is plotted for fixed $\mu = \pi$ and a range of κ-values in Fig. 2G. When κ = 0 the distribution is flat (untuned) and as κ increases the peak is more sharply tuned.

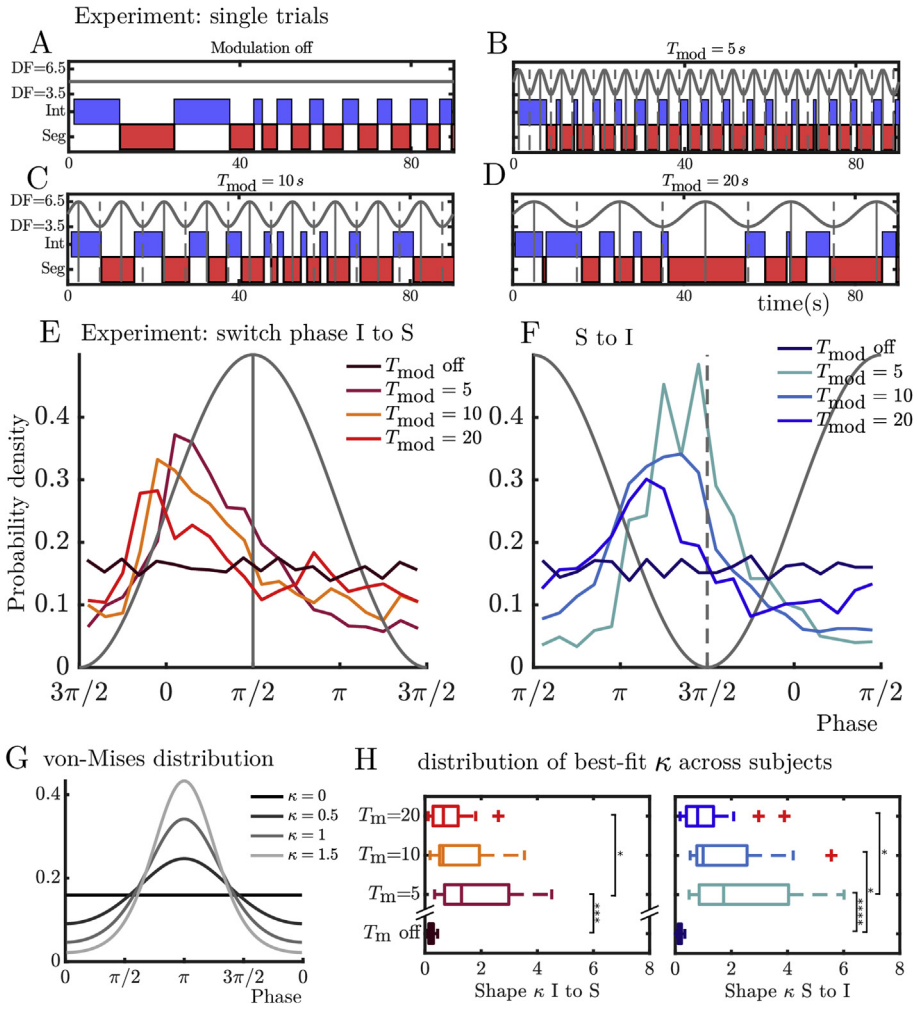

Fig. 2. Experimental confirmation of entrainment. **A-D:** Perceptual responses from individual trials with fixed DF = 5 (**A**, "modulation off" case) or with DF sinusoidally modulated ± 1.5 st with period T_{mod} as indicated (**B**–**D**). **E:** (**F**) Phase histograms averaged across participants ($N = 15$) from Int to Seg (Seg to Int) relative to the modulation of DF centered on the modulation peak (centered on the modulation trough) across 4 T_{mod} conditions as indicated in the legend. **G:** von-Mises distribution plotted for $\mu = \pi$ and different values of the shape parameter κ as used to fit phase hisograms for each participant. **H:** Box and whisker plots showing median, 25th percentile, 75th percentiles and outliers for the distribution across participants of best-fit κ values. Significant differences (Bonferroni corrected) as indicated (otherwise not significant). Note that there is no modulation period for the T_{mod} off case and the *y*-axis is a list of T_{mod} cases rather than a continuous scale.

2.5. Computational model

The neuromechanstic model used here to investigate and predict the effects of periodic feature modulation was presented in Rankin et al. (2015). The model is assumed to be downstream of A1 with three units and pool inputs from three tonotopic locations with best frequencies A, B and (A + B)/2. The dynamic equations describing the time evolution of mean firing rates neural populations (units) pooling inputs that mimic neurophysiological recordings from macaque A1 (Micheyl et al., 2005) are given in Sec. A. Several mechanisms typically found in cortex drive competition between the units: mutual inhibition, spike frequency adaptation and stochasticity (additive noise). Our model is a departure from classical models of perceptual rivalry featuring similar mechanisms but defined in terms of abstracted percepts (Shpiro et al., 2009; Seely and Chow, 2011). Inputs capture the onset-plateau characteristics of A1 responses to pure tones and their dependence on presentation rate and DF. In order to maintain activity during the silent phases between tones and triplets it was necessary to include recurrent excitation on an intermediate NMDA-like timescale. In the present study, the model was adapted to include inputs that depend on time via slow modulation of DF. Several cases were considered in this study: DF fixed at 5 st or sinusoidally modulated about this value with a period $T_{mod} = 5$, 10 or 20 s. The strength of adaptation and noise was increased with respect to the values used in Rankin et al. (2015) in order to match the mean perceptual alternation durations from the no modulation case in the experimental data presented here (values of g and γ described in Sec. A). The model version used here has global inhibition and synaptic depression on the NMDA excitation variables, choices shown to give a better match to behavioral data over a range of DF - values. The equations defining the model are given in Sec. A and a full account of the underlying motivation and assumptions behind the model was given our original paper Rankin et al. (2015).

3. Results

3.1. Prediction from neuromechanistic model

A computational model of auditory bistability (Rankin et al., 2015) was used to predict the perceptual dynamics for the auditory streaming paradigm with slowly modulated DF. For long

stimulus presentations the model accounts for the proportion of time spent integrated (Int, I) or segregated (Seg, S) and the relative length of perceputal durations for static changes to DF as confirmed in behavioral experiments. Here, periodic modulation of DF was introduced noting that the Int percept is more likely at low DF and the Seg percept is more likely at large DF.

A single period of modulation is shown in Fig. 1A for a modulation period $T_{mod} = 10$ s (spanning 20 triplets each of 0.5 s length). The A tones are kept at a fixed frequency and the B frequencies (fixed for a given tone) are modulated about DF = 5 st (where the model produces equidominance, i.e. equal Int and Seg durations). At DF = 5 st the overall mean durations are $T_{eq} = 4.6$ s (with a mean Int duration of 4.2 s and a mean Seg duration of 5.0 s). A single model simulation with DF modulation at $T_{mod} = 10$ s shows apparent 1:1 entrainment (Fig. 1A). Each upswing in DF produces a switch from Int to Seg and each downswing a switch from Seg to Int. Histograms of the stimulus phase of switch times in each direction are shown in Fig. 1C and D. These Phase histograms are centered at the maximal DF value at Phase $= \pi/2$ for switches from Int to Seg (panel C) and minimal DF value at Phase $= 3\pi/2$ for switches from Int to Seg (panel D). Switches from Int to Seg occur on the rising phase of the DF modulation, with a peak close to the steepest part of the curve, when DF crosses through 5 st. Switches from Seg to Int occur on the falling phase with a lower amplitude peak after the steepest decreasing part of the DF modulation. This amplitude asymmetry is likely due to asymmetries captured by the model either side of equidominance; previous work showed that equal semitone decreases of DF gave a stronger bias towards integration compared increases of DF biasing towards segregation (Rankin et al., 2015). The model predicts that alternations entrain to the stimulus modulation with two switches per modulation cycle. Also that switches into segregated (integrated) occur on the DF upswing (downswing), with most switches occurring before the maximal (minimum) DF values. Finally that entrainment is stronger when the DF modulation period is near double the mean percept duration at equidominance, i.e. when $T_{mod} = 2T_{eq}$ (see supplemental material Fig S1). We note that a larger amplitude of DF-modulation was used in the model (2–8 st, compared with 3.5–6.5 st in experiments) because preliminary experiments showed excessively strong entrainment that had been underestimated by the model (discussed further in Sec. 4.4).

3.2. Modulating bistable stimuli entrains perception

For minutes-long presentations of the static auditory streaming paradigm the mean integrated and segregated percept durations (lasting 2–15 s) are roughly equal (equidominant) at DF = 5st and an inter-tone-onset interval of 125 ms (Pressnitzer and Hupé, 2006) or 120 ms (Rankin et al., 2015). Using the same stimulus parameters, in our control condition (example trial Fig. 1A) the overall mean duration was 4.6 s and near equidominance (mean integrated 4.1 s and mean segregated 5.1 s). With DF modulation turned on, the dominance durations can lengthen or shorten, depending on the period of the modulation (Fig. 4B). A repeated measures ANOVA showed a significant effect of T_{mod} on perception durations ($F(3,48) = 8.41$, $p < 0.005$, $\eta_G^2 = 0.09$). For modulation periods roughly equal to twice the modulation off mean dominance duration ($T_{mod} = 10$ s), the switch times become more predictable, with switches from integrated to segregated occurring on the increasing phases of modulation and switches from segregated to integrated occurring on the decreasing phases (Fig. 2C). For faster modulation ($T_{mod} = 5$ s), switching between integrated and segregated become more frequent and more regular (Fig. 2B). For slower modulation ($T_{mod} = 20$ s), the dominance durations on average become longer (Fig. 2D).

Phase histograms showing the probability of switches occurring at different times relative to the modulation of DF are shown in Fig. 2E (Int to Seg) and 2F (Seg to Int). Note that the phase axis represents a longer time window for larger T_{mod} conditions and that the histogram is centered at the peak in modulation (grey curve; Phase $= \pi/2$) in panel E and the trough in modulation (Phase $= 3\pi/2$) in panel F. For all modulation periods, switches from integrated to segregated occurred predominately on the increasing phase of the modulation (Fig. 2E) and switches in the opposite direction occur on the decreasing phase (Fig. 2F). The peak amplitude in the switch phase histogram increases as T_{mod} decreases corresponding to stronger entrainment. Switches from Int to Seg tend to occur earlier with respect to the steepest increasing part of the modulation curve (where DF = 5 st at Phase = 0) compared with corresponding steepest decreasing part for switches from Seg to Int (at Phase $= \pi$). Furthermore, for switches in either direction the peak amplitude is at an earlier phase for larger T_{mod} values.

A circular von-Mises distribution was fit to each of the switch phase histograms in order to assess the strength of entrainment (see Sec. 2). The distribution's shape parameter κ measures how tightly tuned are the data around its centre value, and therefore indicates stronger entrainment for larger values (Fig. 2G). In the modulation off case the phase was arbitrarily defined with a period equal to the mean duration on a trial-by-trial basis. Summary plots of the distribution of best-fit κ values across participants are shown in Fig. 2H, showing no entrainment with modulation off, strongest entrainment for $T_{mod} = 5$ and weaker entrainment for larger T_{mod}. A repeated measures ANOVA test found significant effect of the modulation period on the shape parameter for switches in either direction (I to S $F(3,48) = 8.37$, $p < 0.005$, $\eta_G^2 = 0.24$; S to I $F(3,48) = 14.0$, $p < 10^{-6}$, $\eta_G^2 = 0.30$). A post hoc analysis (pairwise t-test with Bonferroni-corrected significance levels) revealed significant differences between the modulation off condition and the $T_{mod} = 5$ s condition (I to S $p = 0.0003, d = 1.34$; S to I $p = 10^{-4}$, $d = 1.47$) and between the $T_{mod} = 5$ s and $T_{mod} = 20$ s conditions (I to S $p = 0.03, d = 0.89$; S to I $p = 0.04, d = 0.84$) for both directions of switches, and significance between the modulation off and $T_{mod} = 10$ s conditions for segregated to integrated only ($p = 0.04$, $d = 1.01$).

To assess the consistency of the switch phase on a cycle-by-cycle basis, we plotted the probability of reporting the integrated/segregated percept in the cycles before and after a switch of a given phase (Fig. 3). The method for producing these switch-triggered probability maps is discussed in Sec. B and was inspired by modelling work on periodically modulated inputs to sensory neurons (Longtin et al., 1991, 1994). For fast modulation ($T_{mod} = 5$ s), each percept lasts approximately half a cycle and occurs regularly every cycle, as illustrated by the vertically stacked areas of high probability. The horizontally spaced areas of high probability are separated by the length of a triplet, suggesting that participants tend to switch at the same time location within an ABA_ triplet (within-triplet phase). Examining the switch times relative to the triplet phase, in the unmodulated condition, we found that this was in fact the case (see supplemental material Fig S2). Given the transmission and processing delays associated with perceptual switches, we cannot assess which tone in the triplet the switches tend to occur at, only that they occur consistently at the same within-triplet phase. For slower modulation ($T_{mod} = 10$ s and $T_{mod} = 20$ s), we also see horizontally spaced peaks in probability. These peaks are closer together as the modulation periods are longer and as such the triplet length is shorter relatively to the modulation period. Note that the vertically spaced bands, corresponding to successive modulation cycles become increasingly sloped as T_{mod} is increased. Hence, for fast modulation, there is high

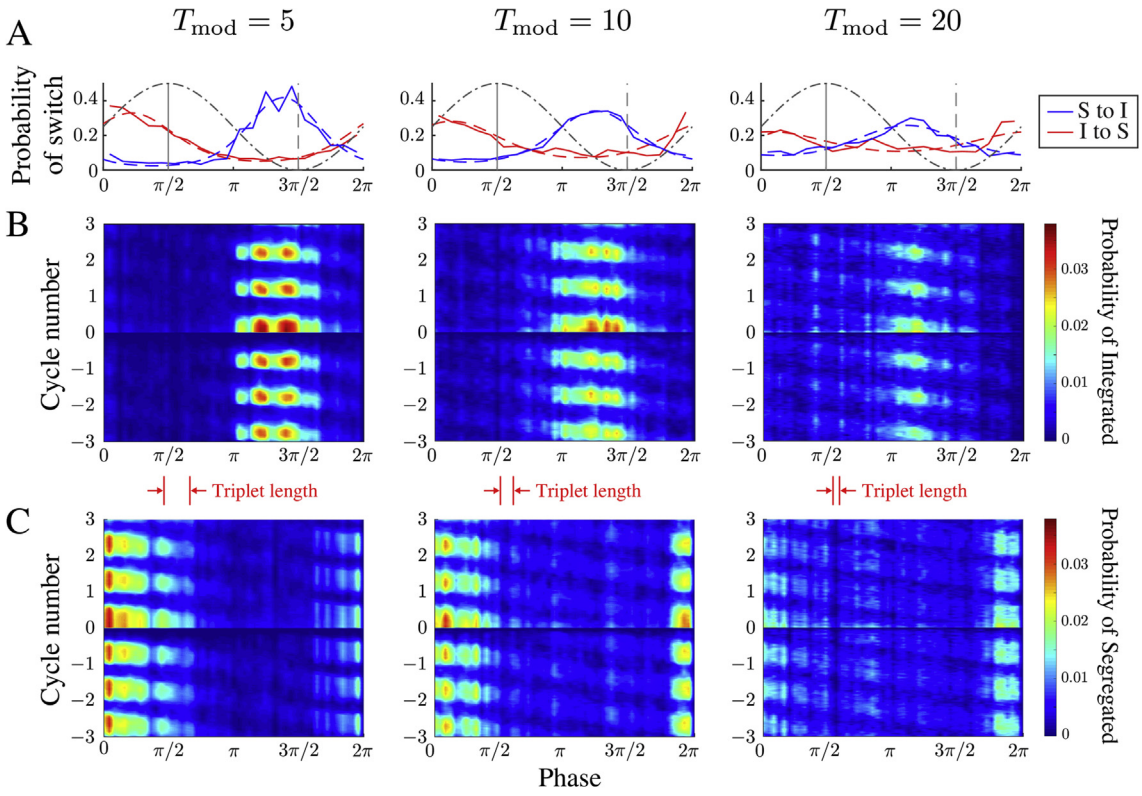

Fig. 3. Characterizing entrainment. **A:** Switching histograms, as in Fig. 2E–F, separated by modulation period. **B–C:** Probability density for each percept (B - integrated, C - segregated) during the modulation cycles before and after a switch of a given phase. The heat maps were computed by examining the durations of each percept before and after every switch and normalizing the probabilities relative to the total number of forcing cycles (see Appendix B for more details). The relative length of a triplet (0.5 s) is shown between each set of plots.

phase consistency. However, for slower modulation, the phase of the switch is less consistent on a cycle-by-cylce basis, i.e. switch that occurs after the mean switch phase is likely to be followed by an earlier phase on the next cycle and vice-versa. The dependence of the preferred phase of a switch on T_{mod} (earlier for large T_{mod}) can also be seen here. For example in Fig. 3C, where the vertical high-probability bands are concentrated before 2π for $T_{mod} = 20$, either side of $0 = 2\pi$ for $T_{mod} = 10$ and above 0 for $T_{mod} = 5$. The presence of multiple peaks, most notably for $T_{mod} = 5$, in the switch histogram curves (Fig. 3A) and multiple vertical bands (panels B and C) is to some extent explained by inter-participant variability as studied further in Sec. 3.3 and supplemental material Figs. S3–S5.

3.3. *Duration histograms suggest different entrainment properties over* T_{mod}

The distribution of percept durations typically follow a log-normal distribution with a peak away from 0 and a long tail (Pressnitzer and Hupé, 2006; Rankin et al., 2015). With DF modulation on, the distribution of durations depends on the modulation frequency (Fig. 4A). For fast modulation ($T_{mod} = 5$ s), there is a tight peak around $T_{mod}/2$ and a very short tail, consistent with strong entrainment. For an intermediate modulation rate ($T_{mod} = 10$ s), the distribution is bimodal, with one peak in a similar location to the peak in the modulation off condition and the other at $T_{mod}/2$. For slow modulation, ($T_{mod} = 20$ s), the peak of the distribution is similar to that of the modulation off case. However, the tail is longer, suggesting some participants may be entraining to the slow modulation, while others ignore the modulation and switch at their natural rate. The mean and variability of the durations is reduced by the DF modulation as illustrated by summary plots of the duration distributions (Fig. 4B). Notably the mean duration for all participants drops from 7.8 s to 4.2 s at $T_{mod} = 5$ s and the coefficient of variation from 0.8 to 0.6 (see values for all conditions in Fig. 4 caption). In the modulation off condition, there is large inter-participant variability with mean switch times ranging between 2 and 30 s (Fig. 4C). In order to explore how different switch rates might interact differently with the modulation period, we grouped the participants as fast ($[\text{mean(I)} + \text{mean(S)}] < 10$), medium ($10 < [\text{mean(I)} + \text{mean(S)}] < 20$) or slow ($[\text{mean(I)} + \text{mean(S)}]/2 > 20$) switchers, depending on their natural switch rate/mean duration in the control experiment.

With the modulation off, fast switchers display a tight distribution (red), while the medium (blue) and slow (green) switchers have similar broad distributions (Fig. 5A T_{mod} off). All three groups show a similar durations distribution for fast modulation, with the medium and slow switchers showing a slightly longer tail (Fig. 5A $T_{mod} = 5$ s). For the intermediate modulation rate ($T_{mod} = 10$ s), the medium and slow switchers show signs of entrainment, a peak at half the modulation period and a short tail (Fig. 5A $T_{mod} = 10$ s). The peak of the fast switchers distribution is in a similar location to that of the modulation off case, however, the distribution has a longer tail, revealing that fast switchers are affected by the modulation to some degree. For slow modulation, the slow switchers have a peak at $T_{mod}/2$, but the fast and medium switchers do not (Fig. 5A $T_{mod} = 20$ s). The tails of the distributions for both the fast and medium switchers are longer than in the modulation off condition. Examining the distribution of mean durations across the three groups, it is clear that modulating DF brings the groups closer together, most notably at $T_{mod} = 5$s (Fig. 5B). Note also that the medium (fast) group's durations are distributed tightly around the

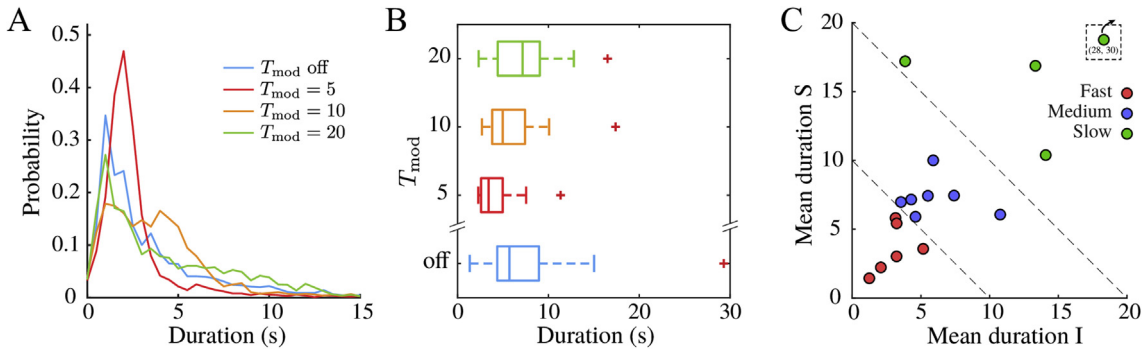

Fig. 4. A: Duration histograms for each TMod condition. Strong peak with short tail suggests strong entrainment for the T_{mod} = 5s condition (red). There appears to be two peaks for the T_{mod} = 10s condition (yellow), suggesting two types of behavior. The peak for the T_{mod} = 20s condition (green) matches the peak for modulation off condition (blue), indicating that participants are not well entrained and operate at their natural switch rate. However, the long tail suggests that some participants may be entraining to the 20 s modulation period. **B:** Box and whisker plots showing the distribution of mean durations across T_{mod} conditions. Modulation reduces the spread of the mean durations, reducing participant-to-participant variability. Means and coefficients of variation (cv) are as follows: T_{mod} off, 7.8 (cv = 0.8); T_{mod} = 5 s, 4.2 (cv = 0.6); T_{mod} = 10 s, 6.0 (cv = 0.6); T_{mod} = 20 s, 7.3 (cv = 0.5). Note that the outlier in each condition corresponds to the same participant. **C:** Scatter plot of mean dwell times in the modulation off condition. Participants were categorized based on mean durations for the modulation off condition; fast (0–5 s), medium (5–10 s) and slow (>10 s). Note that one slow switching participant is off graph (I=28 s, S=30 s). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

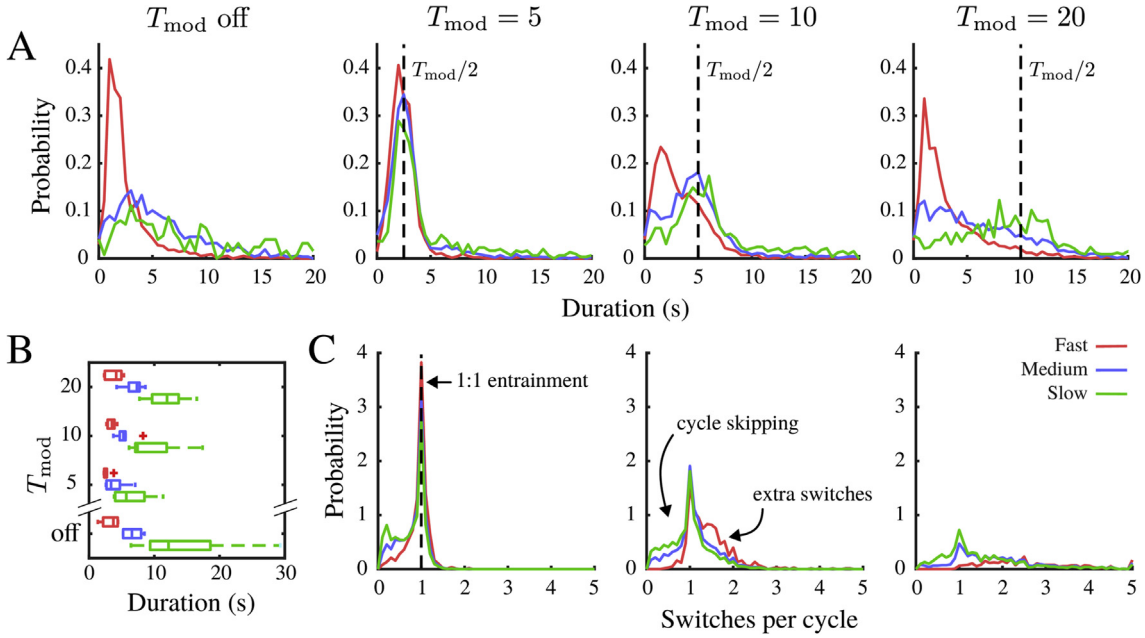

Fig. 5. Participant groupings reveal differences in characteristics of entrainment. **A.** Duration histograms. All participants entrain strongly in the T_{mod} = 5 s condition. Medium and slow participants show characteristics of entrainment at T_{mod} = 10 s, while fast participants have shorter durations. At T_{mod} = 20 s, the slow switchers show a peak around 10s, signifying entrainment. The medium switchers have a longer tail, suggesting some entrainment, whereas the distribution for the fast switchers is almost identical to the unmodulated case. **B:** Box and whisker plots illustrating the distribution of mean durations for each group across T_{mod} conditions. All groups narrow their distributions for T_{mod} = 5 s and T_{mod} = 10 s. However, we do not see a reduction in the spread of the means for the fast and medium switchers when T_{mod} = 20 s. Means and coefficient of variances are given in the supplemental material. **C:** Histograms of the number of switches per modulation period. Sharp peak at 1 for T_{mod} = 5 s condition, confirming strong 1:1 entrainment, with some cycle skipping for the medium and slow switchers. For T_{mod} = 10 s, there is sill a peak at 1 for all 3 groups, but the peak is not as sharp and we see evidence of extra switches for the fast switchers. At T_{mod} = 20 s, the fast switchers no longer have a peak at 1 and show a wide distribution of switches per cycle. The medium and slow switchers peak at 1, but the distributions are much flatter than for the other modulation values. A sliding window analysis was used to calculate the number of switches per cycle, with a window width of 5 T_{mod} and overlap of T_{mod}.

value $T_{mod}/2$for the T_{mod} = 10 s (T_{mod} = 5 s) case (Fig. 5B). In each of these examples, one outlier appears entrained at T_{mod} rather than at $T_{mod}/2$. Using a sliding window analysis, we calculated the number of switches per modulation cycle for each condition and switch group (Fig. 5C). Strong 1:1 entrainment can be seen for T_{mod} = 5 s, with lower peaks at 1 for T_{mod} = 10 s and T_{mod} = 20 s. For T_{mod} = 10 s, there is evidence of cycle skipping for the medium and slow switchers (< 1 switch per cycle) and extra switches for the fast switchers (> 1 switch per cycle). The entrainment at T_{mod} = 20 s is significantly weaker, with much wider distributions about 1. There is no peak at one for the fast switcher, suggesting that either the modulation has no effect on this group or that they are entraining weakly with multiple switches per cycle at 2:1, 3:1, etc. Evidence for extra (multiple) switches per cycle and for cycle skipping across different groups is further elaborated using probability maps (as Fig. 3) in supplemental material Fig. S3–S5.

4. Discussion

Building on research of the ABA_ triplet paradigm for auditory streaming, this work makes a crucial step towards studying dynamic environments. Previous research has typically focused on static environments with fixed stimulus features. One of the key interests in streaming studies stems from the fact that, for a steady repeating stimulus, perception can switch from the initial integrated percept to the segregated percept (build-up of stream segregation). Furthermore, the same fixed stimulus can alternate between integrated and segregated interpretations for long presentations of an unchanging stimulus (auditory bistability). However, natural listening environments are dynamic, sources move and their features evolve over time. An established neuromechanistic model (Rankin et al., 2015) was used to explore perceptual entrainment for a slowly modulated stimulus features. The model predicted entrainment to the modulation with specific phase of switches from integrated to segregated (before the peak in DF) and from segregated to integrated (before the trough in DF). These model predictions served as a springboard for psychoacoustic experiments investigating slow variation of a stimulus feature.

Entrainment of perceptual alternations was confirmed in new experiments, with similar phase to that predicted by the model. We further found that the strength of entrainment and its phase depended on the modulation period. For faster modulation, entrainment was strongest and occurred later in phase. An asymmetry in the phase between the two switch directions (earlier for switches from Int to Seg, Fig. 2E and 2F) could be explained by either the bias (Int 4.1 s and Seg 5.1 s) towards segregation observed in the modulation off case or by previous work showing asymmetries about equidominance where equal semitone decreases of DF gave a stronger bias towards integration compared increases of DF biasing towards segregation (Rankin et al., 2015). The mean of perceptual durations decreased with modulation on, and for the fastest modulation rate appeared to be entrained at the modulation half-period. At slower rates, entrainment was less clear for an all-participants analysis, notably with a bi-modal distribution at an intermediate modulation rate (Fig. 4A). With the aim of teasing apart divergent behaviors across individuals, the participants were split into groups of slow, medium and fast switchers. We found that participants entrained best when the mean of their intrinsic perceptual durations was close to or less than the half-period of the modulation. If the modulation half-period was less than their intrinsic period their alternations entrained, but if the modulation half-period was greater then this value entrainment was much weaker, if present at all.

4.1. Time-varying stimuli

The modulation of stimulus features has been explored previously for auditory streaming. An important aspect of van Noorden's original streaming experiments was the slow increase or decrease of DF that lead to either a switch from integrated to segregated (at the fission boundary) or from segregated to integrated (at the temporal coherence boundary) (van Noorden, 1975). Such unidirectional variation of a stimulus parameter to induce a single perceptual switch was also used in an fMRI study where streams were segregated based on a difference lateralization between two streams (interaural time differences were increased or decreased over 60 s) (Schadwinkel and Gutschalk, 2011). Other studies have considered more complex patterns than the ABA_ triplets used here (Bendixen et al., 2010; Rahne and Sussman, 2009). In these studies stimuli included dynamic changes to additional stimulus features, such as tone intensity, with an unpredictable pattern (not periodic) and without the possibility for entrainment (stimulus presentations were short $\sim$ 10 s).

Auditory streaming is an example of perceptual bistability, which occurs in other sensory modalities including vision (Pressnitzer and Hupé, 2006) and touch (Carter et al., 2008). Conflicting information presented exclusively to each eye leads to spontaneous alternations in perception, so-called binocular rivalry (Lack, 1974; Li et al., 2017; Rodríguez-Martínez and Castillo-Parra, 2018). A computational model and behavioral experiments have investigated the effects of periodic modulation of the difference in contrast between the images presented to each eye (Riani and Simonotto, 1994; Kim et al., 2006). It was found that a modulation period of double the equidominance perceptual duration gave rise to strongest entrainment. In the present study we found equally strong entrainment for modulation periods of less than double the equidominance perceptual duration. Additional peaks at odd multiple of the equidominance half-period were also reported in binocular rivalry experiments (Kim et al., 2006), however, we did not find such peaks in the present study. This may reflect being close to but not exactly at equidominance at DF = 5 st, larger variability in durations for auditory bistability or the inter-participant varaibility discussed below. Recently, the stochastic variation of contrast was considered at different frequencies to probe the internal representation of sensory noise with experiments and a model (Baker and Richard, 2019). A similar approach could be applied with auditory streaming for DF or other stimulus features with an appropriate method for sampling a stochastic process (with a lower bound on the possible modulation frequency range determined by the triplet repetition rate).

4.2. Diverse entrainment across individuals

Individual participants alternate between integration and segregation at different rates (switch durations range from 2 to 15 s, Fig. 4B; see also Rankin et al. (2015), Fig 11B). Differences between individuals have been shown to be stable on a timescale of years (Denham et al., 2014). Inter-participant variability arises for bistable stimuli in other modalities, albeit typically with a smaller range of mean durations, e.g. 2–6 s for binocular rivalry (Patel et al., 2014). Cao et al. (2016) compared the duration statistics across several bistable stimuli including auditory streaming and showed that, whilst most modalities had a coefficient of variation (cv) of around 0.6, durations for auditory bistability are more variable with cv $\approx$ 0.75. Inter-participant variability provides a challenge in auditory bistability experiments, where data needs to be meaningfully combined across individuals; in a minutes-long trial faster switchers may change perception tens of times and slower switchers only once or twice. The modulated stimulus paradigm introduced here can eliminate some of the variability across participants.

At the fast modulation rate, all switchers were similarly entrained (Fig. 5A; column 2 [$T_{mod} = 5$]) to the stimulus. For the intermediate modulation rate medium and slow switchers were entrained (Fig. 5A; columns 3 [$T_{mod} = 10$]) and for the slow modulation rate only the slow switchers were entrained (Fig. 5A; columns 4 [$T_{mod} = 20$]). In general, if the modulation rate is equal to or faster than the intrinsic switch rate of the participant, entrainment was found. From a modelling perspective, the mechanisms that govern intrinsic oscillations (such as inhibition or adaptation) will also govern whether the oscillator tends to entrain when the modulation rate is faster than the intrinsic switch rate. Our experimental results therefore provide a constraint on model design in future studies. We found that mean durations decreased with modulation on at $T_{mod} = 5$ s or at $T_{mod} = 10$ s and that inter-participant variability decreased from a cv = 0.8 with modulation

off to cv = 0.6 or lower for all modulation cases. By grouping participants into slow, medium and fast switchers different entrainment properties were identified between the groups. Such a classification may prove useful in combining data across participants in future studies.

Our analysis of entrainment showed that the phase of switch times was earlier for longer modulation periods (Fig. 2E and 2F). More detailed analysis using probability phase maps revealed peaks in the probability of perceiving a specific percept separated by the length of a triplet (Fig. 3). This suggests that perceptual switches are locked to a specific phase within a triplet, which we confirmed also to be the case for the unmodulated case (supplemental material Fig. S2). Imaging studies have shown that differences in the MEG waveform across percepts arise with a specific latency after the triplet onset (Gutschalk et al., 2005; Billig et al., 2018), which would be consistent with perceptual reports reflecting these differences emerging with specific triplet phase. Further analysis, grouping participants by switch rate, showed that different groups had a tendency to switch around specific triplets, with slower switchers switching later in modulation phase (supplemental material Fig. S3–S5).

4.3. Imaging and neurophysiological experiments

Our results carry practical implications for future auditory bistability experiments with human listeners and animal models. Knowledge of the likely perceptual state as dependent on instantaneous phase of stimulus modulation could allow for the listener to engage in other tasks without the need to report integration or segregation. Their performance in other tasks (say detecting deviants) could be explored with a time-locked a priori expectation of their perceptual state. Furthermore, perceptual reports in imaging experiments can introduce motor artefacts that could be avoided using a modulated paradigm (see recent papers (Costa-Faidella et al., 2017; Billig et al., 2018; Kondo et al., 2018) and a comprehensive review Snyder and Elhilali (2017)). The same expectation could prove useful for animal models where invasive recording is possible, but objective measures of perception are limited, although see Itatani and Klump (2014), Christison-Lagay and Cohen (2014) and Cai et al. (2018).

4.4. Models of auditory streaming

This paper serves a key purpose for computational models: to make predictions that inspire new experiments. Nevertheless, the model used here does not predict every aspect of the experimental data. The strength of entrainment predicted by the model underestimates that found in experiments (model simulations were carried out with a modulation depth of ±3 st, relative to ± 1.5 st in experiments). Preliminary experiments with a modulation depth of ±3 st showed that, with such strong entrainment, it was difficult to discern differences across T_{mod} conditions. Furthermore, while the model predicts a tighter/taller peak in the phase histogram for switches from Int to Seg (Fig. 1C and D), our experiments showed a tighter/taller peak for switches from Seg to Int (Fig. 2E and 2F). Furthermore, our experiments show the strongest entrainment for the fastest modulation (T_{mod}=5 s), whereas the model would predict this at the intermediate modulation rate T_{mod} = 10 s (Fig. S1; although note that T_{mod} = 5 and 10 s show similar peak amplitude). A direct comparison on this point could be misleading as the model represents an *average participant*, where as the experimental data features individuals with different switch rates. A better comparison could be between the medium group (Fig. S4 panel A) and the model (Fig. S1), where a similar amplitude peak is found for T_{mod} = 5 and 10 s with a decreased amplitude for T_{mod} = 20 s in both model and experiments. We further note that the model predicts multiple peaks in the phase histogram for the T_{mod} = 10 s case (a similar effect was observed at T_{mod} = 5 s in the experiments), i.e. that switches occur at a specific within-triplet phase (Fig. S1). The data reported here provides a resource to further constrain models of auditory streaming, including the model used here (Rankin et al., 2015). Indeed a model would provide an excellent test bed for exploring the neural mechanisms that differentiate between individuals with different switching characteristics (like switch rate). Previous studies have reported a balance of adaptation and noise as driving perceptual alternations and the contribution of each to switching dynamics is likely different from individual to individual (Meso et al., 2016). We would predict that stronger entrainment is likely to correlate with more adaptation-driven dynamics, a hypothesis that could be explored in future modelling studies of auditory streaming.

A range of modelling approaches have been proposed for auditory streaming, e.g. based on signal processing (Beauvois and Meddis, 1996), temporal coherence (Krishnan et al., 2014), tonotopic organization (Almonte et al., 2005) or neural oscillations (Wang and Chang, 2008) (recent review: Szabó et al. (2016)). The majority of models have focused on reproducing the dependence of perceptual bias, and/or the dynamics of build-up, on DF and presentation rate (i.e. the van Noorden organization). The present study focused on post-build-up alternations, and recently several models have investigated such auditory bistability with competition dynamics (Mill et al., 2013; Rankin et al., 2015) or probabalistic switching schemes (Steele et al., 2015; Barniv and Nelken, 2015). Indeed, simulations with the model presented in Rankin et al. (2015) (incorporating periodic stimulus modulation) provided the prediction of entrainment (Fig. 1) that inspired the behavioral experiments in the present study. In future work the dataset presented here provides an opportunity to further improve and constrain the model. For example, the range of switch rates of individuals are captured by model parameters governing the strength of internal noise and adaptation (also its timescale) (Rankin et al., 2015). A future study will seek to interpret inter-participant variability and differing interactions with periodically modulated stimuli in terms of cortical mechanisms such as adaptation. Furthermore, other experimental paradigms exploring adaptation dynamics for e.g. sensory memory, feature discrimination thresholds and hearing thresholds could be used to make within-participant predictions for entrainment strength. It is expected that other models based on competition dynamics would also predict entrainment, e.g. Mill et al. (2013) where the competition process is based on similar mechanisms to Rankin et al. (2015) (see also models of binocular rivalry: Riani and Simonotto (1994); Kim et al. (2006)). On the other hand, whilst more abstract models based on probabilistic switching schemes (Steele et al., 2015; Barniv and Nelken, 2015) could be adapted to consider time-varying stimuli, it is not clear how results could be interpreted in terms of dynamics (e.g. predicting the temporal phase of switches).

5. Conclusions

In summary, we found that perception entrains to slowly varying features, that the strength of entrainment depends on the modulation period and that individual differences in a listener's intrinsic switch rate has a marked effect on which stimuli they entrain to. The paradigm presented here is generalizable and could be extended to slow modulation of any stimulus feature over which streams can be segregated or that biases perception towards one percept, e.g. by manipulating localization (Schadwinkel and Gutschalk, 2011), amplitude modulation (Yamagishi et al., 2017) or intensity. Whilst the paradigm presented moves beyond

streaming for static environments, it could also be of broader application for research on streaming. Periodic modulation of a feature enables an investigator to compensate for some of the inter-participant variability found for auditory bistability. Furthermore, with an a priori expectation of the current perceptual state depending on modulation phase, the need for explicit perceptual reports is removed, which could be advantageous in human imaging or animal neurophysiology experiments. Future work should move beyond periodic modulation of stimulus features to consider unpredictable environments that move closer to natural auditory scenes.

Funding

Byrne was funded by the Swartz Foundation on a postdoctoral fellowship. Rankin acknowledges support from an Engineering and Physical Sciences Research Council (EPSRC) New Investigator Award (EP/R03124X/1) and from the EPSRC Centre for Predictive Modelling in Healthcare (EP/N014391/1).

Data availability

All experimental data and model code are available in the github repository james-rankin/auditory-streaming: https://github.com/james-rankin/auditory-streaming.

Declaration of competing interest

Nothing to declare.

A. Model equations

We note two errors in the description of the model in Rankin et al. (2015): inhibition from the r_{AB} unit to the r_A and r_B units is assumed stronger than other inhibitory connections by a factor of 2 (incorrectly reported as a factor of 1 in our earlier study) and the tonotopic decay constant $\sigma_p = 4.25$ (incorrectly reported as twice this value in our earlier study).

The model equations are given by

$$\begin{aligned}
\tau_r \dot{r}_{AB} &= -r_{AB} + F(\beta_e d_{AB} e_{AB} - \beta_i r_{AB} - \beta_i (r_A + r_B) - g a_{AB} + w(\mathrm{DF}(t)/2)(I_A + I_B) + \chi_{AB}),\\
\tau_r \dot{r}_A &= -r_A + F(\beta_e d_A e_A - \beta_i r_A - 2\beta_i r_{AB} - g a_A + I_A + w(\mathrm{DF}(t)) I_B + \chi_A),\\
\tau_r \dot{r}_B &= -r_B + F(\beta_e d_B e_B - \beta_i r_B - 2\beta_i r_{AB} - g a_B + I_B + w(\mathrm{DF}(t)) I_A + \chi_B),\\
\tau_a \dot{a}_{AB} &= -a_{AB} + r_{AB},\\
\tau_a \dot{a}_A &= -a_A + r_A,\\
\tau_a \dot{a}_B &= -a_B + r_B,\\
\tau_e \dot{e}_{AB} &= -e_{AB} + r_{AB},\\
\tau_e \dot{e}_A &= -e_A + r_A,\\
\tau_e \dot{e}_B &= -e_B + r_B,\\
\tau_d \dot{d}_{AB} &= -d_{AB} + (1 - k\, r_{AB}),\\
\tau_d \dot{d}_A &= -d_A + (1 - k\, r_A),\\
\tau_d \dot{d}_B &= -d_B + (1 - k\, r_B).
\end{aligned} \tag{1}$$

The synaptic time constant for each unit is $\tau_r = 10$ ms. The function F translates the synaptic inputs to each population into a firing rate and takes a sigmoidal form

$$F(u) = \frac{1}{1 + \exp(k_F(-u + \theta_F))}, \tag{2}$$

with threshold $\theta_F = 0.2$ and slope $k_F = 12$. Excitation with strength $\beta_e = 0.85$ is assumed to be local, to work on an intermediate NMDA-like timescale of $\tau_e = 70$ ms and undergo slow synaptic depression on a timescale $\tau_d = 3$ s with strength $\kappa = 0.25$. Inhibition with strength $\beta_i = 0.3$ is assumed global and to act instantaneously. Inhibition from the r_{AB} unit to the r_A and r_B units is assumed stronger by a factor of 2 as in Huguet et al. (2014). Spike-frequency adaptation has strength $g = 0.11$ and a slow timescale $\tau_a = 1.4$ s. Inputs to the model mimic the onset-plateau responses to pure tones in A1 with onset timescale $\alpha_1 = 15$ ms, plateau timescale $\alpha_2 = 82.5$ ms and peak to plateau ratio $\Lambda_2 = 1/6$. Inputs are given by the following double α-function where $H(t)$ is the heaviside function.

$$I(t) = H(t)\left[\frac{\exp(2)}{\alpha_1^2} t^2 \exp\left(\frac{2t}{\alpha_1}\right) + \Lambda_2 \frac{\exp(2)}{\alpha_2^2} t^2 \exp\left(\frac{2t}{\alpha_2}\right)\right]. \tag{3}$$

Input amplitudes depend on DF with peak amplitude $I_p = 0.47$, decaying away over tonotopy with a spatial scale of $\sigma_p = 4.25$ st as defined by

$$w(\mathrm{DF}) = I_p \exp\left(\frac{-\mathrm{DF}}{\sigma_p}\right). \tag{4}$$

Additive noise is introduced with independent stochastic processes χ_{AB}, χ_A and χ_B and added to the inputs of each population as in Shpiro et al. (2009) and Seely and Chow (2011). Input noise is modeled as an Ornstien-Uhlenbeck process:

$$\dot{\chi}_k = -\frac{\chi_k}{\tau_d} + \gamma \sqrt{\frac{2}{\tau_X}} \xi_k(t), \tag{5}$$

where $\tau_X = 100$ ms (a standard choice Shpiro et al. (2009); Seely and Chow (2011)) is the timescale, $\gamma = 0.075$ the strength and $\xi(t)$ a white noise process with zero mean. Note these terms appear inside the firing rate function F such that firing rates r_k remain positive and do not exceed 1. Simulations were run in Matlab using a standard Euler-Murayama time stepping scheme with a stepsize of 5 ms (half the value of the fastest timescale in our equations τ_r =

10ms). Reducing this timestep by a factor of 10 did not change the results.

All model code is available in the following GitHub repository: james-rankin/auditory-streaming.

B. Switch triggered probability maps

To create the heat maps in Fig. 3 the perceptual switch times were computed relative to the phase of the modulation. We separated the phase into 50 equally spaced bins, and looked at switches which fell inside each bin. For a given phase bin, the probability of holding a given percept was computed as the number of times a participant held that percept before/after a switch within that bin, divided by the total number of modulation periods. The probability was normalized by the following quantity:

$$\text{Normalizationfactor} = \text{number of participants} \times \text{number of trials} \times \text{trial length} \div T_{\text{mod}}.$$

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heares.2019.107807.

References

Almonte, F., Jirsa, V., Large, E., Tuller, B., 2005. Integration and segregation in auditory streaming. Physica D 212, 137–159.

Anstis, S., Saida, S., 1985. Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. Hum. Percept. Perform. 11, 257–271.

Baker, D.H., Richard, B., 2019. Dynamic properties of internal noise probed by modulating binocular rivalry. PLoS Comput. Biol. 15, 1–18. https://doi.org/10.1371/journal.pcbi.1007071.

Barniv, D., Nelken, I., 2015. Auditory streaming as an online classification process with evidence accumulation. PLoS One 10 e0144788.

Beauvois, M., Meddis, R., 1996. Computer simulation of auditory stream segregation in alternating-tone sequences. J. Acoust. Soc. Am. 99, 2270–2280.

Bee, M., Klump, G., 2004. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J. Neurophysiol. 92, 1088–1104.

Bee, M., Micheyl, C., Oxenham, A., Klump, G., 2010. Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. J. Comp. Physiol. 196, 543–557. https://doi.org/10.1007/s00359-010-0542-4.

Bendixen, A., Denham, S.L., Gyimesi, K., Winkler, I., 2010. Regular patterns stabilize auditory streams. J. Acoust. Soc. Am. 128, 3658–3666.

Billig, A.J., Davis, M.H., Carlyon, R.P., 2018. Neural decoding of bistable sounds reveals an effect of intention on perceptual organization. J. Neurosci. 38 (11), 2844–2853. https://doi.org/10.1523/JNEUROSCI.3022-17.2018.

Cai, H., Screven, L.A., Dent, M.L., 2018. Behavioral measurements of auditory streaming and build-up by budgerigars (melopsittacus undulatus). J. Acoust. Soc. Am. 144, 1508–1516.

Cao, R., Pastukhov, A., Mattia, M., Braun, J., 2016. Collective activity of many bistable assemblies reproduces characteristic dynamics of multistable perception. J. Neurosci. 36, 6957–6972. https://doi.org/10.1523/JNEUROSCI.4626-15.2016.

Carter, O., Konkle, T., Wang, Q., Hayward, V., Moore, C., 2008. Tactile rivalry demonstrated with an ambiguous apparent-motion quartet. Curr. Biol. 18, 1050–1054.

Christison-Lagay, K.L., Cohen, Y.E., 2014. Behavioral correlates of auditory streaming in rhesus macaques. Hear. Res. 309, 17–25.

Costa-Faidella, J., Sussman, E.S., Escera, C., 2017. Selective entrainment of brain oscillations drives auditory perceptual organization. Neuroimage 159, 195–206.

Denham, S.L., Bohm, T., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill, R., Winkler, I., 2014. Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. Front. Neurosci. 8, 1–15.

Denham, S.L., Farkas, D., Van Ee, R., Taranu, M., Kocsis, Z., Wimmer, M., Carmel, D., Winkler, I., 2018. Similar but separate systems underlie perceptual bistability in vision and audition. Sci. Rep. 8.

Farley, B.J., Noreña, A.J., 2015. Membrane potential dynamics of populations of cortical neurons during auditory streaming. J. Neurophysiol. 114, 2418–2430. https://doi.org/10.1152/jn.00545.2015. http://jn.physiology.org/content/114/4/2418.

Fishman, Y., Arezzo, J., Steinschneider, M., 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. 116, 1656–1670.

Fishman, Y., Reser, D., Arezzo, J., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167–187.

Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., Oxenham, A.J., 2005. Neuromagnetic correlates of streaming in human auditory cortex. J. Neurosci. 25, 5382–5388.

Huguet, G., Rinzel, J., Hupé, J.M., 2014. Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. J. Vis. 14, 19.

Itatani, N., Klump, G., 2014. Neural correlates of auditory streaming in an objective behavioral task. Proc. Natl. Acad. Sci. U.S.A. 111, 10738–10743.

Kashino, M., Kondo, H., 2012. Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. Philos. Trans. R. Soc. Lond. B 367, 977–987.

Kim, Y.J., Grabowecky, M., Suzuki, S., 2006. Stochastic resonance in binocular rivalry. Vis. Res. 46, 392–406.

Knyazeva, S., Selezneva, E., Gorkin, A., Aggelopoulos, N.C., Brosch, M., 2018. Neuronal correlates of auditory streaming in monkey auditory cortex for tone sequences without spectral differences. Front. Integr. Neurosci. 12, 4.

Kondo, H.M., Kashino, M., 2009. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. J. Neurosci. 29, 12695–12701.

Kondo, H.M., Pressnitzer, D., Shimada, Y., Kochiyama, T., Kashino, M., 2018. Inhibition-excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability. Sci. Rep. 8, 14548.

Krishnan, L., Elhilali, M., Shamma, S., 2014. Segregating complex sound sources through temporal coherence. PLoS Comput. Biol. 10 e1003985.

Lack, L.C., 1974. Selective attention and the control of binocular rivalry. Percept. Psychophys. 15, 193–200.

Li, H.H., Rankin, J., Rinzel, J., Carrasco, M., Heeger, D., 2017. Attention model of binocular rivalry. Proc. Natl. Acad. Sci. U.S.A. 114 (30), E6192–E6201. https://doi.org/10.1073/pnas.1620475114.

Longtin, A., Bulsara, A., Moss, F., 1991. Time-interval sequences in bistable systems and the noise-induced transmission of information by sensory neurons. Phys. Rev. Lett. 67, 656.

Longtin, A., Bulsara, A., Pierson, D., Moss, F., 1994. Bistability and the dynamics of periodically forced sensory neurons. Biol. Cybern. 70, 569–578.

Meso, A.I., Rankin, J., Faugeras, O., Kornprobst, P., Masson, G.S., 2016. The relative contribution of noise and adaptation to competition during tri-stable motion perception. J. Vis. 16 https://doi.org/10.1167/16.15.6, 6–6. http://jov.arvojournals.org/article.aspx?articleid=2593028.

Micheyl, C., Tian, B., Carlyon, R., Rauschecker, J., 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139–148.

Mill, R., Bőhm, T., Bendixen, A., Winkler, I., Denham, S., 2013. Modelling the emergence and dynamics of perceptual organisation in auditory streaming. PLoS Comput. Biol. 9 e1002925.

van Noorden, L., 1975. Temporal Coherence in the Perception of Tone Sequences. PhD Thesis. Eindhoven University.

Patel, V., Stuit, S., Blake, R., 2014. Individual differences in the temporal dynamics of binocular rivalry and stimulus rivalry. Psychon. Bull. Rev. 1–7.

Pressnitzer, D., Hupé, J., 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr. Biol. 16, 1351–1357.

Pressnitzer, D., Sayles, M., Micheyl, C., Winter, I., 2008. Perceptual organization of sound begins in the auditory periphery. Curr. Biol. 18, 1124–1128.

Rahne, T., Sussman, E., 2009. Neural representations of auditory input accommodate to the context in a dynamically changing acoustic environment. Eur. J. Neurosci. 29, 205–211.

Rankin, J., Osborn Popp, P., Rinzel, J., 2017. Stimulus pauses and perturbations differentially delay or promote the segregation of auditory objects: psychoacoustics and modeling. Front. Neurosci. 11 https://doi.org/10.3389/fnins.2017.00198.

Rankin, J., Sussman, E., Rinzel, J., 2015. Neuromechanistic model of auditory bistability. PLoS Comput. Biol. 11 https://doi.org/10.1371/journal.pcbi.1004555 e1004555.

Riani, M., Simonotto, E., 1994. Stochastic resonance in the perceptual interpretation of ambiguous figures: a neural network model. Phys. Rev. Lett. 72, 3120.

Rodríguez-Martínez, G.A., Castillo-Parra, H., 2018. Bistable perception: neural bases and usefulness in psychological research. Int. J. Psychol. Res. 11, 63–76.

Schadwinkel, S., Gutschalk, A., 2011. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. J. Neurophysiol. 105, 1977–1983.

Seely, J., Chow, C.C., 2011. Role of mutual inhibition in binocular rivalry. J. Neurophysiol. 106, 2136–2150.

Shpiro, A., Moreno-Bote, R., Rubin, N., Rinzel, J., 2009. Balance between noise and adaptation in competition models of perceptual bistability. J. Comput. Neurosci. 27, 37–54.

Snyder, J.S., Elhilali, M., 2017. Recent advances in exploring the neural underpinnings of auditory scene perception. Ann. N. Y. Acad. Sci. https://doi.org/10.1111/nyas.13317.

Steele, S., Tranchina, D., Rinzel, J., 2015. An alternating renewal process describes the buildup of perceptual segregation. Front. Comput. Neurosci. 8, 1–13.
Szabó, B.T., Denham, S.L., Winkler, I., 2016. Computational models of auditory scene analysis: a review. Front. Neurosci. 10 https://doi.org/10.3389/fnins.2016.00524.
Wang, D., Chang, P., 2008. An oscillatory correlation model of auditory streaming. Cogn Neurodyn. 2, 7–19.
Winkler, I., Denham, S.L., Mill, R., Bőhm, T.M., Bendixen, A., 2012. Multistability in auditory stream segregation: a predictive coding view. Philos. Trans. R. Soc. Biol. Sci. 367, 1001–1012.
Yamagishi, S., Otsuka, S., Furukawa, S., Kashino, M., 2017. Comparison of perceptual properties of auditory streaming between spectral and amplitude modulation domains. Hear. Res. 350, 244–250.